

The North Atlantic Population Project

An Overview

EVAN ROBERTS
STEVEN RUGGLES
*Minnesota Population Center
University of Minnesota*

LISA Y. DILLON
*Département de Démographie
Université de Montréal*

ÓLÖF GARDARSDÓTTIR
Statistics Iceland

JAN OLDERVOLL
*Department of History
University of Bergen*

GUNNAR THORVALDSEN
*Norwegian Historical Data Centre
University of Tromsø*

MATTHEW WOOLLARD
*History Data Service
University of Essex*

Abstract. The North Atlantic Population Project (NAPP) brings together complete-count census data from late-nineteenth-century Canada, Great Britain, Iceland, Norway, and the United States into a single harmonized database. When released in 2005, the final version of the database will include the records of nearly 90 million people. The project will consistently code all variables across the different countries, while still retaining important national variation in census questions and responses. The authors provide a brief history of the project, discuss the main issues involved in creating a harmonized international census database, and outline the methodological and research opportunities the completed database will provide for scholars.

Keywords: census, harmonization, microdata, North Atlantic Population Project (NAPP)

The North Atlantic Population Project (NAPP) brings together late-nineteenth-century complete-count census microdata from Canada, Great Britain, Iceland, Norway, and the United States into a single harmonized database.¹ When finished in 2005, the database will contain information on nearly 90 million individuals from these five countries. All samples will be available in a common format with consistent variable coding and careful documentation of the differences among the censuses. Preliminary release of the data began in December 2001 with individual country data sets, and we are planning the first release of the harmonized data set for summer 2003.²

This article provides a brief history of the project, a discussion of the main issues involved in creating the database, and an overview of methodological and research opportunities the completed database will provide scholars. Occupational classification is the most complex and time-consuming aspect of the project and has received much of our attention so far. The major issues in classifying occupations

in the NAPP database are discussed in a companion article by Roberts et al. (pp. 89–96 in Part Two of this issue).

Background

The NAPP will incorporate complete, machine-readable versions of late-nineteenth-century population censuses of Canada, Great Britain, Iceland, Norway, and the United States. At the close of the nineteenth century, these countries were closely connected by substantial flows of people and commerce. Each nation was undergoing profound social, economic, and demographic changes, including fertility and mortality decline, increasing income and migration, and growing economic complexity. A harmonized database comprising these entire North Atlantic populations will enhance our understanding of these connections and transitions.

The NAPP is made possible by the availability of complete machine-readable transcriptions of nineteenth-century census enumerators' manuscripts. The Church of Jesus Christ of Latter-day Saints (LDS), in collaboration with local genealogical societies, laboriously transcribed and entered into databases the 1881 censuses for Canada and Great Britain and the 1880 census of the United States, to provide a resource for genealogical research. Thousands of professionals and volunteers contributed an estimated total of 11 million hours of work entering and verifying information on the 85 million individuals who lived in these three countries at the beginning of the 1880s.

In Norway and Iceland, researchers have created databases to serve as scientific resources. Over the past four decades, Norwegian researchers and archivists have invested more than half a million hours digitizing historical population records. The national censuses of 1865 and 1900 are

complete, and the census of 1875 is fully digitized for one-quarter of the population, and a 2 percent sample covers the rest. The Norwegian database is a collaborative product of the Norwegian Historical Data Centre at the University of Tromsø, the Digital Archive of the Norwegian National Archives, and the University of Bergen. Statistics Iceland computerized the oldest nominal Icelandic census dating from 1703, and Daniel Vasey led a project in the early 1990s to digitize the censuses of 1801, 1845, and 1870.³ More recently, researchers have transcribed information from the censuses of 1880, 1901, and 1930 in an effort to construct genealogies for genetic research.⁴

In 2000, participants from each country met in Minneapolis to define the goals of a harmonization project and develop a detailed plan of work. The participants agreed that we should not simply create compatible data sets but rather should develop a single fully harmonized database with common coding systems, constructed variables, documentation, and dissemination systems. In collaboration with U.K., Norwegian, Icelandic, and Canadian scholars, the Minnesota Population Center (MPC) submitted a proposal to the National Science Foundation in early 2001, and the project officially commenced in July 2001 (Grant SES 0111707).⁵

Collaborators

The NAPP is a collaborative project composed of research teams from each of the five countries included in the database. These five countries are fortunate to have extraordinarily rich collections of surviving individual-level census data, as shown in table 1. The United States has the Integrated Public Use Microdata Series (IPUMS), provid-

ing samples of all surviving censuses from 1850 to 2000. A new project in Canada to create twentieth-century samples will fill out a comparable series for 1871 to 2001. Great Britain, Norway, and Iceland have multiple samples or complete-count data sets for the late nineteenth century, surviving manuscripts for the early twentieth century, and existing samples for the late twentieth century. We envision the NAPP as the foundation for a long-term collaborative enterprise to reconstruct the population of this region from the mid-nineteenth century to the present.

The NAPP database will include the seven censuses marked as “N” in table 1: it will combine the three LDS-transcribed censuses of Great Britain, Canada, and the United States with two censuses each from Iceland and Norway.⁶ National research teams have prepared—or are in the process of preparing—these data for public use. The NAPP collaborators have pieced together funding for data preparation from numerous sponsors to support the painstaking tasks of data cleaning and coding. This project does not fund these basic data preparation activities; rather, it is intended to coordinate those activities to produce a harmonized database.

The NAPP involves researchers from the Universities of Montréal and Ottawa in Canada, the University of Essex in Great Britain, Statistics Iceland, the Universities of Bergen and Tromsø in Norway, and the University of Minnesota in the United States. Each of the research teams brings extensive experience with large-scale nineteenth-century census data; most of the participants are key researchers on projects to create microdata samples and complete-count data sets listed in table 1.⁷ Coordinating work in five countries on two continents is challenging. We communicate extensively through Web-based tools, e-mail, telephone, and annual

TABLE 1. Availability of North Atlantic Census Microdata, 1835–2001

	1840		1860		1880		1900		1920		1940		1960		1980		2000	
Great Britain	M	S	M	M	N	M	P	M		M		M	M	Z	Z	S	Z	
Canada		Z	M	S	N	Z	S	Z	Z	Z	Z	M	S	S	S	S	Z	
Iceland	C	M	C	M	C	N	P	P	N	P	M	C	M	M		S		
Norway					N	P	M	N	M	M	M		M	M	S	S	S	Z
United States			S	S	S	N		S	S	S	Z	S	S	S	S	S	S	Z

Key

- N = complete machine-readable data sets to be included in the NAPP database
- M = manuscript individual-level census survives
- S = machine-readable national sample of census exists
- C = complete machine-readable transcription exists
- P = complete machine-readable census transcription planned or in progress
- Z = machine-readable national census sample planned

Note. In Great Britain and Canada, the decennial censuses were carried out in the “1” years (1851, 1861, etc.). U.S. censuses were carried out in the “0” years (1850, 1860, etc.). No manuscripts survive for the 1921 census of Great Britain. No census was held in Great Britain in 1941 because of World War II. For 1951 and 1961, the manuscript returns were punched onto computer cards and converted to a computer-tape format.

meetings of all collaborators. The first two meetings—held in Minneapolis in November 2001 and in Grundarfjörður, Iceland, in October 2002—focused on developing common coding schemes for occupations and other key variables. It can also be difficult to reach agreement when a large group is concerned, but the collaboration is guided by the philosophy that decisions should be reached by consensus and if that is not possible, then a majority decision is required.

Harmonizing the NAPP Database

Harmonization of these North Atlantic censuses is eased because of a high degree of comparability in the original enumeration processes across countries. This comparability was due primarily to the formal as well as informal collaboration of North American and European statisticians and to the development of international standards in forums such

as the International Statistical Congress and the International Statistical Institute. For example, Joseph Kennedy, the United States census superintendent, became a national statistical expert in part through his interactions with European statisticians (Anderson 1988, 46, 60). Enumeration practices in the United States as well as across the Atlantic informed the design of the nineteenth-century Canadian censuses (Curtis 1994, 2000; Worton 1998).

Table 2 describes the data files to be included in the North Atlantic database. With respect to their structure, organization, and available information, the seven censuses are remarkably comparable. Each census enumerated individuals, and in each case, those individuals are grouped into households, which were defined in similar terms in each country. All countries defined a *household* as “a group of people sharing a common place of residence.” There is a core set of variables common to virtually all data sets,

TABLE 2. Variables in the NAPP Database

Country	Great Britain	Canada	Iceland		Norway		United States
Census year	1881	1881	1870	1901	1865	1900	1880
Enumeration rule	de facto	de jure	both	both	de jure	both	de jure
No. of person records (000)	29,000	4,280	70	79	1,702	2,240	50,155
<i>Household record</i>							
Household characteristics							
County	X						X
City, town, village	X	X	X	X	X	X	X
Province/state		X	X	X	X	X	X
Parish	X		X	X	X		
Enumeration district	X	X	X	X		X	X
School district					X		
Address	X		X	X	X	X	
Microfilm reel or folio number	X	X	X	X	X	X	X
Census page number	X	X	X	X	X	X	X
Number and type of rooms						X	
Farm residence	C	C	X	X	X	X	C
Constructed household variables							
Record type (household)	C	C	C	C	C	C	C
Household sequence number	C	C	C	C	C	C	C
Number of persons in household	C	C	C	C	C	C	C
Group-quarters residence	C	C	C	C	C	C	C
Group-quarters type	C	C	C	C	C	C	C
Urban/rural residence	C	C	C	C	C	C	C
Size of place	C	C	C	C	C	C	C
Metropolitan area	C	C	C	C	C	C	C
Community characteristics	C	C	C	C	C	C	C
Household type (UN system)	C	C	C	C	C	C	C
Household type (Hammel/Laslett)	C	C	C	C	C	C	C
Number of families	C	C	C	C	C	C	C
Size of primary family	C	C	C	C	C	C	C
Number of children under 18	C	C	C	C	C	C	C
Number of married couples	C	C	C	C	C	C	C
Number of secondary individuals	C	C	C	C	C	C	C

(table continues)

TABLE 2. (Continued)

Country	Great	Canada	Iceland		Norway		United
	Britain						States
Census year	1881	1881	1870	1901	1865	1900	1880
Enumeration rule	de facto	de jure	both	both	de jure	both	de jure
No. of person records (000)	29,000	4,280	70	79	1,702	2,240	50,155
<i>Person record</i>							
Individual characteristics							
Relationship to household head	X	C	X	X	X	X	X
Age	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X
Occupation	X	X	X	X	X	X	X
Marital status	X	X	X	X	X	X	X
Place of birth	X	X		X	X	X	X
Parental birthplaces							X
Citizenship/nationality	X				X	X	
Ethnicity/race		X			X	X	X
Religion		X			X	X	
Disability	X				X	X	
Surname	X	X	X	X	X	X	X
Given name	X	X	X	X	X	X	X
Absent or visiting on census day			X	X		X	
Constructed person variables							
Record type (person)	C	C	C	C	C	C	C
Person number in household	C	C	C	C	C	C	C
Socioeconomic scores	C	C	C	C	C	C	C
Surname similarity code	C	C	C	C	C	C	C
Location of spouse	C	C	C	C	C	C	C
Location of own mother	C	C	C	C	C	C	C
Location of own father	C	C	C	C	C	C	C
Number of own children	C	C	C	C	C	C	C
Number of own children under 5	C	C	C	C	C	C	C
Age of eldest own child	C	C	C	C	C	C	C
Age of youngest own child	C	C	C	C	C	C	C

Note. X = variable taken directly from source. C = constructed variable.

including relationship of each individual to the household head, age, sex, marital status, occupation, and birthplace. Geographical precision for place of residence varies across countries, but it is possible to identify all places with populations over 5,000 for each country. Across the five countries, we identify approximately 25,000 communities. These common core variables will allow us to construct a variety of new variables describing community and neighborhood characteristics, household composition, socioeconomic status, and family interrelationships.

Most of the censuses were taken on a de jure basis, under which individuals who were temporarily absent from home—such as migrant workers and travelers—were to be enumerated at their usual place of residence. The exception is the British census, taken under a de facto rule, which specified that no one who was present on census night at a particular address could be left out of the tally and that no person absent from home could be written in. From 1870 onward, persons in Norway and Iceland were to be enumer-

ated both at their usual place of residence *and* at the place they stayed on census day, and enumerators identified both temporary visitors and absent household members. These variations in enumeration rules pose only minor compatibility problems, but investigators of some household composition and migration issues will have to be aware of them.

There are two main tasks necessary to create the harmonized NAPP database. First, we must develop common classification systems for occupation, birthplace, family relationship, and group-quarters type that balance the goals of international comparability, retention of detail, and ease of use. Second, we must build a consistent set of constructed variables describing household composition, family interrelationships, urban and metropolitan residence, and other geographic characteristics.

The most challenging task is the development of harmonized coding systems to allow comparison across countries. At present, much of the data are in the form of alphabetic character strings, representing the transcription from the

original source material. These strings are too unwieldy for statistical analysis; to make the data usable, we must develop numeric codes for each variable. For variables such as age, sex, marital status, and religion, this process is straightforward, because the variation in values for these variables across countries is not great. Other variables are more complicated; for example, we estimate that there are unique strings for 2 million occupations, 1 million birthplaces, and 30,000 family relationships.

To translate from character strings into numeric codes, we must construct data dictionaries that assign a numeric code to each alphabetic variation that occurs in the data. This work is difficult enough in the context of a single country; for a project of this scale, it requires a team of expert coders working in close cooperation, sharing coding decisions continuously. We have created a set of merged dictionaries of unprecedented scale that include the alphabetic strings from 90 million persons in all five countries. This is uncharted territory. Until recently, constructing such a dictionary would have necessitated assembling experts with appropriate language skills and historical knowledge about each country in a single location, a prohibitively expensive task. Using Internet technology, however, we are able to distribute the task, thus working on the dictionaries in multiple countries simultaneously. For a description of this coding process for occupation, the most complicated variable, see the Roberts et al. companion article (pp. 89–96 in Part Two of this issue).

When feasible, the coding is compatible with the IPUMS-USA and IPUMS-International coding systems. We also follow the same IPUMS principle of retaining detail while providing harmonized codes suitable for comparative research. Composite coding systems allow us to provide information available across all censuses in a least-common-denominator code and information unique to one country (or set of countries) in a subsequent “detail” code. We also preserve country-specific variables with information or coding systems unique to a particular country; these include, for example, religion (only available in Canada and Norway) and occupations coded to the 1881 British classification scheme. This sensitivity to the particular conventions of each country’s data is important for maintaining productive collaboration and will maximize the usefulness of the finished database. In addition, researchers will be able to download the raw data and analyze, replicate, or alter the coding systems applied by the NAPP team.

We will create a number of constructed variables to facilitate the analysis of household composition, family interrelationships, and own-child fertility. These household- and person-level variables are detailed in table 2. Among the most important of these are pointer variables that indicate the location within each household of each individual’s own spouse, mother, and father, as well as number of own children.⁸ These pointer variables allow researchers to attach the characteristics of immediate kin (spouse’s birthplace, father’s occupation, children’s characteristics, etc.) without

the need to resort to programming (Ruggles 1995). We will add the variables required for basic own-child fertility analysis: number of own children present, number under 5 years of age, age of eldest child, and age of youngest child. We also plan to construct a set of variables to describe the composition of families and households, replicating the most commonly used historical and contemporary classification systems, as well as household characteristics such as the number of persons, families, children, married couples, and servants in each household. Finally, we will create technical variables to aid in data management and analysis, such as record type, serial number, and group-quarters residence.

Documenting and Disseminating the NAPP Database

In addition to data preparation, the project involves documenting the comparability of census enumeration instructions, procedures, and definitions across the five countries, creating machine-understandable metadata compliant with the Data Documentation Initiative (DDI) metadata standard, and developing efficient Web-based software to optimize and simplify access to the entire database.

The creation of integrated documentation is central to the project, and the development of these materials will be a collaborative enterprise. We are working on comprehensive documentation for each of the censuses in the database, including facsimiles of census forms and enumerator instructions. When the original documentation is in another language, we plan to translate the most essential material into English. When foreign-language material is extensive, however, we will provide the full text in the original language with a brief English-language translation. The documentation system will also describe all procedures undertaken to generate the harmonized database, including the actual computer code, data dictionaries, and a textual description of the data manipulation process. All documentation will be compliant with the DDI metadata standard (see Block and Thomas, pp. 97–101 in Part Two of this issue).

Few users will be interested in analyzing the entire data set of 90 million cases. Therefore, we will disseminate the database through a data extraction system that integrates access to metadata and microdata and allows users to carry out substantial manipulation of the data. The data access system for the North Atlantic database will be based on a similar system currently under development as part of the IPUMS-International project (see Ruggles et al., pp. 60–65 in Part Two of this issue). Web sites in Great Britain, Canada, Norway, and the United States will mirror the NAPP data access system and provide unrestricted access for academic researchers.

Applications for Complete-Count Data

The availability of information on entire populations will open important new avenues of research. The following dis-

discussion describes some new methodological approaches that will be possible with complete-count data.

Study of small dispersed population subgroups. Many small population subgroups—defined by race, ethnicity, occupation, or even age—can be studied only with very high density data sets. For example, the availability of complete-count data will allow the study of the indigenous populations of Canada, Norway, and the United States. The 1881 Canadian census microdata include over 60,000 aboriginal persons; these cases would include sufficient indigenous women of childbearing age to allow in-depth fertility analysis. Similarly, existing sample data are inadequate to study immigrant groups in detail. A substantial proportion of the Icelandic population emigrated to Canada and the United States, but there are insufficient Icelandic cases in any of the existing samples to allow quantitative analysis. This is also true for the Irish population in England. The number of Norwegians in North America is larger, but there are still not enough for detailed analysis using a sample database. The new database will also allow, for the first time, comparative study of particular age groups such as centenarians. It will also have adequate cases to compare specific occupations (e.g., sailors, fishermen, and sons of farmers) across all five nations.

Community studies. The community study has been one of the most fruitful analytical approaches in both history and sociology. Most sample data lack sufficient cases to enable researchers to examine particular localities. Because the new database includes the entire population, it will allow scholars to extract customized data sets focusing on particular communities. The international dimension of the database will allow investigators to undertake comparative community studies. For example, researchers could compare patterns in a Minnesota Norwegian community with the sending community in Norway. Investigators could also examine cross-border Canadian-U.S. or English-Scottish communities, to compare economic and social patterns in neighboring areas distinguished only by political and institutional structures. There is a large demand for local historical statistical data, and the North Atlantic database will immediately become an essential tool for community historians of all sorts. Detailed personal identifiers for individuals—including full name, age, sex, occupation, marital status, and birthplace—create the potential for researchers to link NAPP data with other contemporaneous sources of basic demographic information, such as ship lists, school and college attendance records, friendly society (benefit-society) membership lists, as well as birth, death, marriage, and emigration registers. Even historians who make little use of quantitative analysis will be able to quickly and painlessly locate their study subjects in the manuscript census.

Longitudinal analysis. Perhaps the greatest limitation of the existing samples is that they are cross-sectional snapshots

and do not allow one to trace individuals across time. This problem will be alleviated by the new database. As shown in table 1, for each country there are machine-readable samples or complete-count data sets for multiple years. Thus, it will be possible to create a series of linked samples; in the case of Canada, for example, individuals in the 1871, 1891, and 1901 census samples can be linked with the complete-count 1881 census. Thus, researchers will be able to construct three linked Canadian samples, covering 1871–1881, 1881–1891, and 1881–1901. In Norway, Iceland, and Great Britain—where there are multiple complete censuses—there is potential to link individuals across more than a single pair of census years. Researchers will even be able to link some individuals across countries; for example, from Norway in 1865 to Canada in 1881 and back to Norway in 1900.

Linked census data hold the promise of finally resolving some of the longest-running debates in nineteenth-century social history. Past studies of social and geographic mobility were not entirely conclusive because of their exclusion of migrants and their small sample size. Scholars will be able to gauge the extent of social and geographic mobility, analyze the interrelationship of geographic and economic movement, and assess trends and differentials in social mobility far more reliably than heretofore (Thorvaldsen 1995). In addition, the linked samples will allow investigation of questions regarding family formation and dissolution. For example, they will allow scholars to answer several controversial questions surrounding the formation of multigenerational households in the nineteenth century (Ruggles 1994, 2001).

Multilevel analysis. In recent years, analyses of the effects of local context on individual behavior have proven exceedingly valuable tools for research in historical sociology (see Elman 1998; Kramarow 1995; Ruggles 1997a, 1997b). New methods of multilevel analysis will increase the effectiveness of this approach (DiPrete and Forrestal 1994). A key problem for such nineteenth-century research, however, is that the method requires independent variables tabulated for small geographic units, and such data are scarce before the twentieth century. The new North Atlantic database will allow creation of a wide variety of contextual variables—such as racial or ethnic composition, female labor-force participation, and occupational structure—at many geographic levels, including the neighborhood and the enumeration district.

Geographic information systems. Geographers are ordinarily unable to tap the power of microdata. The existing nineteenth-century microdata files are samples, so when they are used for small areas they provide insufficient precision for reliable mapping. Although some relatively high density samples are available for the period since 1970, those microdata files suppress detailed geographic data. Therefore, geographers are forced to rely on complete-count aggregate data that usually provide only basic summary statistics for small areas.

The North Atlantic census database will provide full geographic detail for every individual in the population. Digitized small-area boundary files are already available for Great Britain in 1881 and are in preparation for nineteenth-century Canada, Norway, and the United States (Gregory and Southall 1998; Thorvaldsen 1997; see also Fitch and Ruggles, pp. 41–51 in Part One of this issue).⁹ Therefore, a large scholarly investment in nineteenth-century geographic information systems already exists. What is lacking is a fine level of geographic detail in social, economic, and demographic characteristics. The North Atlantic census database will allow scholars to marry existing geographic boundary files to population characteristics, thus creating a powerful new analytic tool. Such fine geographic analysis will be especially potent in the analysis of topics such as early suburban development and racial and ethnic residential segregation (see, for example, Gardner 1998).

Substantive research areas. A cross-section encompassing the entire population of the North Atlantic world in the late nineteenth century will address many of the major themes of history, economics, demography, and sociology. The following four examples describe some of these key substantive issues:

1. The North Atlantic database will allow in-depth analysis of the processes of industrialization. The first Industrial Revolution may have begun in rural England, but by the late nineteenth century the entire North Atlantic world was involved in manufacturing, production of raw materials, or both. The North Atlantic database will allow unprecedented opportunities to explore economic structures within and between each nation during this critical transitional period. For the first time, we will have consistently coded individual-level occupational data available for nineteenth-century censuses from multiple countries. This will allow comparative analysis at the level of persons, families, communities, or regions, as well as allow investigation of the geographic organization of economic activity. Thus, the NAPP database will allow scholars to address international and North American regional differences in the scale and location of manufacturing activity, comparing, for example, the distribution of particular industrial workers in large urban centers versus smaller towns and villages.

2. The database will also contribute to understanding of the fertility transition. At the time these censuses were taken, each of the North Atlantic countries was witnessing the impact of widespread deliberate fertility limitation.¹⁰ The North Atlantic database will allow study of differential fertility patterns in this critical period of demographic transition, to assess the importance of such factors as occupational class, ethnicity, region, literacy, local economy, size of locality, and family structures. Past comparative analyses of the European fertility transition have relied on national-, regional-, or parish-level aggregate vital statistics (Coale and Watkins 1986). This approach has two major disadvan-

tages. First, aggregate vital statistics do not allow direct measures of child spacing or stopping behavior; only the *level* of fertility can be considered. Second, the aggregate approach does not allow control of individual-level socioeconomic characteristics. Study of this elemental shift in population structure has the potential to enhance our understanding of ongoing demographic change in the contemporary developing world (Woods 2000).

3. Research on the family will also benefit from the common set of constructed variables in the NAPP database. These variables aid in the analysis of family and household composition and will thus allow consistent comparisons across all five countries. It will enable investigators to assess the impact of local context on family systems through multilevel analysis and so for the first time permit analysis of the effects of individual-level factors, local economic conditions, regional inheritance systems, and national characteristics on the nineteenth-century family.

4. Finally, the North Atlantic database will constitute an outstanding resource for the study of migration history. It will allow close and consistent comparisons of occupational structure, marriage patterns, fertility, and family composition. Researchers will be able to identify and compare specific sending and receiving communities. In some instances, it will even be possible to follow individual migrants across the Atlantic and back again. The NAPP database will also facilitate the study of cross-border populations.¹¹ In combination with new machine-readable ship lists and emigration registers, the database will open a new window on the implications of international population flows.

Conclusion

This article has outlined the ongoing project to create the North Atlantic Population Project database. Capitalizing on three decades of experience in creating census samples, the NAPP applies them to a complete-count database including data from five nations. Harmonizing five databases of this scale poses a range of methodological and practical challenges. The potential payoff, however, is great; we are confident that the twin benefits of complete-count data and the ability to compare countries will have substantial benefits for social science research. The NAPP will be compatible with existing and forthcoming time-series of census data for all those countries contributing data. We hope it will spur future efforts to increase the sample density of existing census samples and promote comparative and transnational research.

NOTES

1. For this article, *Great Britain* is taken to mean Great Britain (i.e., England, Wales, and Scotland), the Channel Islands, and the Isle of Man. (On the same date as the rest of the United Kingdom, a separate census was taken in Ireland, the returns of which were destroyed.)

2. Preliminary raw-format data and documentation are currently available at <http://www.nappdata.org>. Future releases will be available through

an online extraction system based on the IPUMS and IPUMS-International systems.

3. On the Icelandic Population Project (IDP), see Vasey 1997.

4. See <http://www.decodegenetics.com/products/ga/> [accessed: 21 September 2002].

5. Funding for the initial data cleaning and coding has come from a wide variety of sources. In Great Britain, the sources of funding include the Economic and Social Research Council, the Leverhulme Trust, and the University of Essex Research Promotion Fund; in Canada, the Social Sciences and Humanities Research Council, the Harold Crabtree Foundation, the Church of Jesus Christ of Latter-day Saints (LDS), and the University of Ottawa Research Partnerships Programme; in Norway, the Norwegian Research Council, the Norwegian National Archives, and the Faculty of Social Sciences of the University of Tromsø; and in the United States, the National Science Foundation and the National Institutes of Health.

6. Additional complete censuses will be included as time and funding permit. Large portions of nineteenth-century Danish censuses are computerized; the entire 1890 census for Sweden was recently made available. See <http://www.foark.umu.se/census/> [accessed: 22 October 2002]. The National Archives of Canada has recently launched a project to digitize at least the first 10 questions of the complete 1901 Census of Canada. There are plans to digitize the whole 1910 census for Norway by 2010. The English and Welsh census for 1901 has also been made machine-readable for genealogical research and is available over the Internet for a fee. See <http://www.census.pro.gov.uk> [retrieved: 17 October 2002].

7. Lisa Dillon was involved in the creation of national census samples of the United States in 1850 and Canada in 1901 and harmonized the 1870 and 1900 United States census samples with the Canadian samples for 1871 and 1901 (Dillon 1997, 2000). Chad Gaffield is currently involved in the Canadian Century Research Infrastructure project to create new samples of Canada for the period 1911 to 1951. Jan Oldervoll and Gunnar Thorvaldsen have worked for two decades to create the most comprehensive electronic archive of historical census data anywhere. Steven Ruggles at the MPC has directed projects to create samples of the U.S. census for the period 1850 through 1930, and he is principal investigator for the IPUMS projects to harmonize U.S. and international censuses. Matthew Woollard, working with Kevin Schürer, has had long experience with the 1851 and 1881 censuses of Great Britain and has extensive knowledge of occupational classification schemes. Ólöf Garðarsdóttir, head of population and vital statistics at Statistics Iceland, has extensive experience with Icelandic and other Nordic historical demographic data. Evan Roberts, the project coordinator, comes from New Zealand.

8. The Canadian census did not enumerate relationship to household head, a key variable for the study of household composition, family inter-relationships, and own-child fertility measures. Accordingly, we will construct an inferred family-relationship variable for the Canadian database, drawing on procedures developed to construct relationship to household head for the 1850, 1860, and 1870 IPUMS sample databases (Ruggles and Sobek 1998; Dillon 1997, 2000). These procedures use surname similarity, sex, age, and position in the household, among other variables, to infer and impute household relationships. Previous efforts to construct a relationship-to-household-head variable for the 1871 Canadian census database have demonstrated the usefulness of marital status, religion, and ethnicity in predicting links among household members; as a result, imputation procedures conducted on the Canadian data will make use of these additional variables. To facilitate the construction of relationship to household head as well as pointer variables in the Canadian data, the Université de Montréal NAPP team is undertaking the cleaning of last names. In the 1881 Canadian census database, French Canadian names were particularly poorly transcribed by the typically Anglophone LDS volunteers. Expertise in French-language name cleaning developed at the Université de Montréal to create the *Registre de la Population de Québec Ancien* is now being applied to the 1881 Canadian census microdata to help resolve such problems.

9. In Canada, historical geographic information projects are under way for late-nineteenth-century Montréal, Québec City, and Toronto; these projects will incorporate the 1881 census microdata for these cities and will also map the 1881 census districts for Québec (Olson 2002; St.-Hilaire et al. 2002; Ashby 2002). Historical geographic information systems will also be developed by the Canadian Century Research Infrastructure project to map early-twentieth-century census data (Gaffield et al. 2001). We anticipate that these initiatives will lead to future projects to geocode the whole of 1881 Canada.

10. Until recently, demographers thought that American fertility decline began much earlier than in most of Northern Europe. J. D. Hacker (1999, 2000), however, has shown by means of census-based own-child and back-projection methods that deliberate fertility limitation in the United States began considerably later than was previously thought. Hacker places the fertility transition in America after the Civil War, the same period when it occurred in each of the other countries in the North Atlantic database.

11. The 1881 Canadian census microdata include over 10,000 American-born persons, whereas the 1880 U.S. census microdata features 706,843 Canadian-born individuals. Similarly, in 1881, there were over 250,000 Scottish natives in England and Wales and 90,000 English or Welsh who had been born in Scotland.

REFERENCES

- Anderson, M. J. 1988. *The American census: A social history*. New Haven: Yale University Press.
- Ashby, D. 2002. Personal e-mail communication to Lisa Dillon. Toronto: Department of Geography and Planning, University of Toronto. 9 March.
- Coale, A. J., and S. C. Watkins. 1986. *The decline of fertility in Europe: The revised proceedings of a conference on the Princeton European fertility project*. Princeton: Princeton University Press.
- Curtis, B. 1994. On the local construction of statistical knowledge: Making up the 1861 census of the Canadas. *Journal of Historical Sociology* 7: 416–34.
- . 2000. *The politics of population: State formation, statistics, and the census of Canada, 1840–1875*. Toronto: University of Toronto Press.
- Dillon, L.Y. 1997. Between generations and across borders: Living arrangements of the elderly and their children in Victorian Canada and the United States. Ph.D. diss., University of Minnesota.
- . 2000. Integrating Canadian and U.S. historical census microdata. *Historical Methods* 33: 185–94.
- DiPrete, T. A., and J. D. Forristal. 1994. Multilevel models: Methods and substance. *Annual Review of Sociology* 20: 331–57.
- Elman, C. 1998. Intergenerational household structure and economic change at the turn of the twentieth century. *Journal of Family History* 4: 417–40.
- Ferrie, J. P. 1999. *Yankees now: Immigrants in the antebellum United States, 1840–1860*. New York: Oxford University Press.
- Gaffield C. et al. 2001. Canadian Century Research Infrastructure (CCRI). Funding Application to the Canadian Foundation for Innovation, Project # 5291. Ottawa: Institute of Canadian Studies, University of Ottawa. May.
- . 2002. Des données de recensement historiques sur Internet grâce à des recherches sans précédent. *Initiatives*. Ottawa: Institute of Canadian Studies: 1–2.
- Garðarsdóttir, Ó. 2002. *Saving the child: Regional, cultural and social aspects of the infant mortality decline in Iceland, 1770–1920*. Umeå: Demographic Data Base, Umeå University.
- Gardner, T. 1998. The metropolitan fringe: Suburbanization in the United States before World War II. Ph.D. diss., University of Minnesota.
- Garrett, E. et al. 2001. *Changing family size in England and Wales: Place, class and demography, 1891–1911*. Cambridge: Cambridge University Press.
- Gregory, I. N., and H. R. Southall. 1998. Putting the past in its place: The Great Britain historical GIS. In *Innovations in GIS 5*, edited by S. Carver, 210–21. London: Taylor & Francis.
- Guest, A. 1987. Notes from the National Panel Study: Linkage and migration in the late nineteenth century. *Historical Methods* 20: 63–77.
- Gunnlaugsson, G. Á., and Ó. Garðarsdóttir. 1996. Transition into widowhood: Life course perspective on the household position of Icelandic widows at the turn of the century. *Continuity and Change* 11: 435–58.
- Hacker, J. D. 1999. The human cost of war: White population in the United States, 1850–1880. Ph.D. diss., University of Minnesota.
- . 2000. Rethinking the “early” decline of American fertility. Paper presented at the Minnesota Population Seminar, October.
- Hareven, T. K. 1982. *Family time and industrial time: The relationship between the family and work in a New England industrial community*. New York: Cambridge University Press.
- Katz, M. B. 1975. *The people of Hamilton, Canada West: Family and class in a mid-nineteenth-century city*. Cambridge: Harvard University Press.
- Knights, P. R. 1991. *Yankee destinies: The lives of ordinary nineteenth-century Bostonians*. Chapel Hill: University of North Carolina Press.

- Kramarow, E. 1995. The elderly who live alone in the United States: Historical perspectives on household change. *Demography* 32: 335–52.
- McCalla, D., and K. Inwood. 2002. The public use microsample of the 1891 census of Canada for Ontario: A project overview. Guelph: Department of History, University of Guelph. <<http://www.uoguelph.ca/history/census/index.htm>> [created: 21 March 2002] [retrieved: 21 September 2002].
- Olson, S. 2002. MAP: Montréal l'avenir du passé (Géoïde project HSS#56). Sharing geodatabases in historical dynamics. *Annual Report*. Montréal: McGill University.
- Ruggles, S. 1994. The origins of African-American family structure. *American Sociological Review* 59: 136–51.
- . 1995. Family interrelationships. *Historical Methods* 28: 52–58.
- . 1997a. The rise of divorce and separation in the United States, 1880–1980. *Demography* 34: 455–66.
- . 1997b. The effects of AFDC on American family structure, 1940–1990. *Journal of Family History* 22: 307–25.
- . 2001. Living arrangements and economic well-being of the aged in the past. *Population Bulletin of the United Nations* 42/43: 111–61.
- Ruggles, S., and M. Sobek. With C. A. Fitch, P. K. Hall, and C. Ronnander. 1998. *Integrated Public Use Microdata Series: Version 2.0., vol. 1: User's Guide*, edited by C. A. Fitch. Minneapolis: Historical Census Project, University of Minnesota.
- St.-Hilaire, M. et. al. 2002. Population et histoire sociale de la ville de Québec. Laval: Laboratoire de géographie historique, Université Laval. <<http://www.phsvq.cieq.ulaval.ca/>> [retrieved: 21 September 2002].
- Thernstrom, S. 1964. *Poverty and progress: Social mobility in a nineteenth-century city*. Cambridge: Harvard University Press.
- Thorvaldsen, G. 1995. *Migration in the province of Tromsø, 1865–1900. Migrasjon i Troms i annen halvdel av 1800-tallet. En kvantitativ analyse av folketellingene 1865, 1875 og 1900*. Diss.—from the University of Tromsø—with summary in English, <<http://www.rhd.uit.no/nhdc/summary.html>>.
- . 1997. Om grenser og områder i lokalhistorisk forskning / On borders and areas in local history research. In *Festskrift til Sivert Langholm*, edited by J. Myhre and K. Kjeldstadli, 188–211. Oslo: Norwegian Historical Association. <http://www.rhd.uit.no/census/boundaries.html> [retrieved: 21 September 2002].
- Vasey, D. E. 1997. An estimate of neonatal tetanus mortality in Iceland, 1790–1839. *European Journal of Population* 13: 49–69.
- Woods, R. 2000. *The demography of Victorian England and Wales*. Cambridge: Cambridge University Press.
- Worton, D. A. 1998. *The Dominion Bureau of Statistics: A history of Canada's Central Statistical Office and its antecedents, 1841–1972*. Montréal: McGill-Queen's University Press.

OUR ARCHIVES

now available online.

Powerful keyword search.
Easy to use.
Find articles quickly.

www.heldref.org

Historical Methods is published by Heldref Publications,
a division of the nonprofit Helen Dwight Reid Educational Foundation

